4 Computation

Technically, for a given value of $\boldsymbol{\theta}$, define $L_n(\boldsymbol{\theta}|\boldsymbol{Y}) = 0$ if

$$\mathcal{A}_n(\boldsymbol{\theta}) = \left\{ \prod_{i=1}^{n} p_i : p_i \mapsto \boldsymbol{Y}_i, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) = \mathbf{0}_r \right\}$$

is empty. (EL function might not be computable over all possible parameter $\underline{\theta}$ values).

If $\mathcal{A}_n(\underline{\theta})$ is empty, $\mathcal{A}_n(\underline{\theta}) = \emptyset$, then $L_n(\underline{\theta})$ is not defined

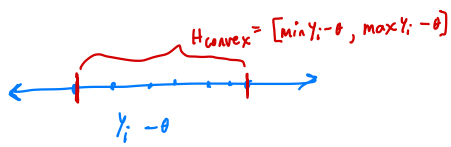
$\Rightarrow$ Define $L_n(\underline{\theta}) = 0$ (this is the smallest is can be anyways).

If $\mathbf{0}_r$ is in the $\underbrace{\text{interior convex hull of } \{\psi(\mathbf{Y}_i, \boldsymbol{\theta})\}_{i=1}^{n}}_{H_{convex}}$, then $\mathcal{A}_n(\boldsymbol{\theta})$ will not be empty. $\Rightarrow L_n(\underline{\theta}) > 0$.

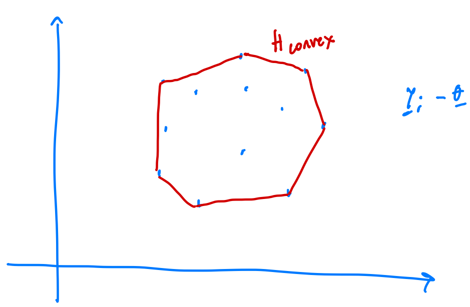
H_{convex} is the smallest convex set containing $\psi(\underline{Y}_1, \underline{\theta}), \dots, \psi(\underline{Y}_n, \underline{\theta})$.

E.g. if $\psi(\underline{Y}_i, \underline{\theta}) = \underline{Y}_i - \underline{\theta}$ (sample mean case), and

$q = 1$:



$q = 2$:



The supremum in the definition of $L_n(\boldsymbol{\theta}|\boldsymbol{Y})$ looks nasty, but the form simplifies if $L_n(\boldsymbol{\theta}|\boldsymbol{Y}) > 0$ for a given $\boldsymbol{\theta} \in \mathbb{R}^b$. To see this, fix $\boldsymbol{\theta}$ and let

interior of.

If $\underline{\theta}$ is in H_{convex}

$$\mathcal{B}_n(\boldsymbol{\theta}) = \left\{(p_1, \ldots, p_n) : p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) = \mathbf{0}_r \right\} \subset [0,1]^n$$

closed and bounded (compact) set in $\mathbb{R}^n$

$\Rightarrow$ the supremum $L_n(\underline{\theta}) = \sup \mathcal{A}_n(\underline{\theta}) = \sup\{\prod_{i=1}^n p_i : (p_1, \ldots, p_n) \in \mathcal{B}_n(\underline{\theta})\}$ is attained as a maximum of $\prod_{i=1}^n p_i$ on $\mathcal{B}_n(\underline{\theta})$. So $\exists$ $p_1^*, \ldots, p_n^* \in \mathcal{B}_n(\underline{\theta})$ where $L_n(\underline{\theta}) = \prod_{i=1}^n p_i^*$.

(Uniqueness) Suppose $q_1^*, \ldots, q_n^* > 0$ and $p_1^*, \ldots, p_n^* > 0$ in $\mathcal{B}_n(\underline{\theta})$ such that $\prod_{i=1}^n p_i^* = \prod_{i=1}^n q_i^*$.

Now let $r_i^* = \alpha p_i^* + (1-\alpha) q_i^*$ for $\alpha \in [0,1]$.

If $\alpha \in (0,1)$, it holds that $\sum_{i=1}^n \log r_i^* > \alpha \sum_{i=1}^n \log p_i^* + (1-\alpha) \sum_{i=1}^n \log q_i^*$ (Jensen's).

If $\prod_{i=1}^n q_i^* = \prod_{i=1}^n p_i^* = M$ holds, where $M > 0$ is the maximum of $\prod_{i=1}^n p_i$ on $\mathcal{B}_n(\underline{\theta})$, then

$\sum_{i=1}^n \log r_i^* > \sum_{i=1}^n \log p_i^*$, which cannot be true because $\prod_{i=1}^n p_i^*$ is the max of $\prod_{i=1}^n p_i$ on $\mathcal{B}_n(\underline{\theta})$.

$\Rightarrow$ $(p_1^*, \ldots, p_n^*)$ which maximize $\prod_{i=1}^n p_i$ on $\mathcal{B}_n(\underline{\theta})$ must be unique!

To maximize $\prod_{i=1}^{n} p_i$ on $\mathcal{B}_n(\boldsymbol{\theta})$ and find $(p_1^*, \ldots, p_n^*)$, use Lagrange multipliers $a \in \mathbb{R}$ and $\boldsymbol{\lambda} \in \mathbb{R}^r$ and maximize

$$f(p_1, \ldots, p_n, a, \boldsymbol{\lambda}) = \log \prod_{i=1}^{n} p_i + a \underbrace{\left(1 - \sum_{i=1}^{n} p_i\right)}_{\text{probability constraint}} - n\boldsymbol{\lambda}^\top \underbrace{\left(\sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta})\right)}_{\text{EE moment constraint.}}$$

over $p_i \in [0,1]$, $a \in \mathbb{R}$, and $\boldsymbol{\lambda} \in \mathbb{R}^r$.

Take derivatives & set to zero:

$$\frac{\partial}{\partial p_i} f(p_1,\ldots,p_n,a,\underline{\lambda}) = \frac{1}{p_i} - a - n\underline{\lambda}^T\underline{\psi}(\underline{y}_i,\underline{\theta}) \overset{\text{set}}{=} 0 \implies \boxed{a p_i = 1 - n p_i \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})}$$

$$\sum_{i=1}^{n} a p_i = \sum_{i=1}^{n}\{1 - n p_i \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})\}$$

$$a\sum_{i=1}^{n} p_i = n - n\underline{\lambda}^T\sum_{i=1}^{n} p_i \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$\frac{\partial}{\partial a} f(p_1,\ldots,p_n,a,\underline{\lambda}) = 1 - \sum_{i=1}^{n} p_i \overset{\text{set}}{=} 0 \implies a = n - n\underline{\lambda}^T\sum_{i=1}^{n} p_i \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$\frac{\partial}{\partial \underline{\lambda}} f(p_1,\ldots,p_n,a,\lambda) = -n\sum_{i=1}^{n} p_i \underline{\psi}(\underline{y}_i,\underline{\theta}) \overset{\text{set}}{=} \underline{0} \implies \underline{\underline{a = n.}}$$

$$a p_i = 1 - n p_i \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$p_i = \frac{1}{a} - \frac{1}{a} n p_i \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$p_i = \frac{1}{n} - p_i \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$p_i\left(1 + \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})\right) = \frac{1}{n}$$

$$p_i = \frac{1}{n}\left(\frac{1}{1 + \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})}\right)$$

So, if $\underline{\theta} \in H_{\text{convex}}$, then

$$L_n(\underline{\theta}) = \prod_{i=1}^{n} \frac{1}{n}\left(\frac{1}{1 + \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})}\right)$$ where $\underline{\lambda}$ is determined by solving.

$$\underline{0} = \sum_{i=1}^{n} p_i \underline{\psi}(\underline{y}_i,\underline{\theta})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\underline{\psi}(\underline{y}_i,\underline{\theta})}{1 + \underline{\lambda}^T \underline{\psi}(\underline{y}_i,\underline{\theta})}.$$

See Art Owen's website for code / R package.